DOCUMENT RESUME

ED 170 625 CG 013 813

AUTHOR Thornton, George C., III
TITLE Psychometric Properties of Self-Appraisals of Job Performance.
PUB DATE Aug 78
NOTE 21p.; Paper presented at the Annual Convention of the American Psychological Association (Toronto, Ontario, Canada, August, 1978)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Rating; Behavior Rating Scales; Employees; *Job Satisfaction; *Performance Criteria; *Personnel Evaluation; *Psychometrics; *Self Evaluation; Supervisory Activities; *Task Performance; Work Attitudes

ABSTRACT

The research literature on the psychometric properties of self-appraisals of work performance summarizes many results in terms of leniency, variability, halo, and construct validity. Comparisons with appraisals tend to show more leniency, less variability, more bias, and less discriminant validity. Different factor structures are found among self- and supervisor-ratings, while self-appraisals show generally less halo and significant correlations with other sources in some studies. Existing data do not allow any firm conclusions as to whether the quality of self-appraisals is a function of scale format, amount of rater training, type of judgment, or purpose of appraisal. Problems may exist when self-appraisals are used for administrative decision making, diagnosis of training needs, applied criterion measurement, measurement of constructs in basic research, or for selection purposes. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED170625

Psychometric Properties of Self-Appraisals Of Job Performance

George C. Thornton, III
Colorado State University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G C Thornton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Running Head: Self-appraisals

In E. L. Levine, Organizational Applications of Self-Appraisals and Self-Assessment: Another Look. Presented at the meeting of the American Psychological Association, Toronto, Ontario, Canada, August, 1978.

CG012413

Psychometric Properties of Self-Appraisals Of Job Performance

George C. Thornton, III
Colorado State University

Abstract

This paper reviews the research literature on the psychometric properties of self-appraisals of work performance. It summarizes results with regard to leniency, variability, halo, and construct validity. Comparisons with appraisals by supervisors, peers, and subordinates suggest that self-appraisals tend to show more leniency, less variability, more bias, and less discriminant validity. Different factor structures have been found among self-and supervisor-ratings. On the other hand, self appraisals show generally less halo and significant correlations with other sources in some studies. Existing data do not allow any firm conclusions whether the quality of self-appraisals is a function of scale format, amount of rater training, type of judgment, or purpose of appraisal. The effects of the observed psychometric qualities of self-appraisals on various applications are discussed. Problems may exist when they are used for administrative decision making, diagnosis of training needs, applied criterion measurement, measurement of constructs in basic research, or for selection purposes. Potential benefits are also noted.

Introduction

The rationale for obtaining employee's appraisal of their own job performance has been articulated by McGregor (1957 & 1959) as one part of a more general participative approach to management. Subsequent research by Meyer and his colleagues at General Electric (Bassett & Meyer, 1968; Meyer, Kay & French, 1965) demonstrated that self appraisal and goal setting reduced defensiveness in the performance appraisal interview and led to improved performance on the job. Concerns about the adequacy of self-appraisals were raised in a series of studies by researchers at Case-Western Reserve University. Prien and Liske (1962) and Parker, Taylor, Barrett, & Martens (1959) found that self appraisals showed more leniency and less variability than supervisory appraisals. Thornton (1966 & 1968) found overrating among top executives in a corporate-wide performance appraisal system. Lawler (1967) found more convergent and discriminant validity for supervisor and peer appraisals than for self appraisals. He suggested that self ratings should be used for research purposes to provide a unique perspective on job performance. For example, instrumentality theories of work motivation hypothesize that a variety of job behaviors, including effort and performance are a function of perceptions of the linkage of effort and performance and performance and outcomes. Campbell and Pritchard

(1976) argued that measures of the individual's cognitive performance should be used in expectancy theories. Two studies by Mendel and Dickinson (1971; undated) found that expectancy theory in fact predicts cognitive performance (i.e. self-ratings) better than objective performance.

In this and other literature we can see that self-appraisals have been used for a variety of purposes. It will be helpful in reviewing the research evidence and considering its implications to articulate these at the beginning:

(1) Evaluation of current performance for administrative purposes,

(2) Identification of training needs, e.g. evaluation of job skills and knowledge, career planning,

(3) Criterion measurement in applied research, e.g. test validation, evaluation of training,

(4) Reporting past achievements at time of application for a new job, e.g. background interviews, bio-data forms, the job sampling approach to merit systems examination developed by Schwartz (1977).

(5) Measurement of contructs in basic research, e.g. measurement of self-perceived performance and effort in studies of motivation theories.

As we shall see later, the psychometric properties of self-appraisals have differential relevance for these various applications.

Method

This paper is a review of the literature on self-appraisals of job performance. To be included in the review, the research must have met the following criteria:

(1) data based, using quantitative appraisals,

(2) comparison of self-appraisals with appraisals from other sources,

(3) the subjects must be organizational employees or supervisors, and

(4) self-appraisals of on-the-job performance.

In other words, I was interested in studies which allowed an evaluation of self-appraisals of actual job performance. I found 19 studies.

The following psychometric properties of self-appraisals were reviewed:

* Leniency - the extent to which self-appraisals are higher than appraisals from other sources.

* Variation - the variability or range of ratings for a rater group.

* Halo - intercorrelation among several rating scales, i.e. "individual halo".

* Reliability - stability over time.

* Bias - lack of correlation of self-ratings and ratings from other sources, "point of view".

* Discriminant validity - ability to meet criteria set forth by Campbell & Fiske (1959) in which reliability is high and method bias is low.

* Factor Structure

Results

Leniency

The preponderance of studies show that individuals rate themselves higher than they are rated by some comparison groups. These findings hold for several levels of employees including technical employees (Kirchner, 1966, Klimoski & London, 1974), supervisors (Waldman & Thornton, 1978), and executives (Thornton, 1968). Over-rating appeared for trait and behavioral ratings (Kirchner, 1966) and for job-related and person-related dimensions (Parker et al, 1959). Over-rating increased with time, i.e. in second year of an operational program (Thornton, 1966). It was more pronounced among older employees dissatisfied with pay (Lawler, 1968), among low performers (Thornton, 1968), especially low performers with high self-esteem (Baird, 1977). Individuals may be aware that their supervisors will give lower ratings, but estimates of supervisors' ratings are still significantly higher than supervisor-ratings, falling about half-way between the two (Parker et al, 1959).

The exception to the general finding is a study Heneman (1974) in which self-ratings were significantly lower for three of nine dimensions. The author speculates that the lower self-ratings reflect data collection under a research condition. More studies are needed on the effect of instructional set because all other previous research was also under research conditions, except Thornton's (1968) study of an operational program for executives. Another explanation for Heneman's findings may be sample bias among the respondents of this mail survey.

Seventy-five percent of the managers and 55.5% of their supervisors responded. It might be that the responding supervisors were favorably disposed toward the individuals whereas supervisors of low performers were among the non-respondents.

Variation

Fewer studies report variance in ratings. In the five published accounts, four (Hall, 1976; Klimoski & London, 1974; Parker et al, 1959; Prien & Liske, 1962) found less variation among the self-appraisals. Again, Heneman (1974) found the opposite. Sample biases from his mail survey may have truncated the supervisors' ratings.

Halo

Where several scales are rated by the subject it is possible to examine their interrelationships. The literature contained eleven studies where comparisons of the inter-scale correlations were possible. In general, these correlations were high, a finding that points up the general problem to be discussed later - lack of discriminant validity. In nine studies (Hall, 1976; Heneman, 1974; Kirchner, 1966; Klimoski & London, 1974; Lawler, 1968; Mitchell & Albright, Parker et al, 1959; Porter & Lawler, 1968; Thornton, 1966; Williams & Seiler, 1973) of the eleven studies, self-appraisals manifested less halo than ratings by the comparison group. In other words, self-appraisals showed relatively low inter-scale correlations. Baird (1977) and Nealey and Owen (1970) found no difference in halo, although

correlations for both groups were high in both studies.

At first glance, the high leniency and low variance would seem to make the low halo a statistical artifact. This is not an adequate explanation because the scale correlations for both sources are often in the 50's, 60's, and 70's.

Bias

In the context of this paper, bias means lack of correlation between self-ratings and ratings by others. The findings are inconsistent. Ten studies reported lack of agreement with at least one other source: the immediate supervisor (Baird, 1977; Charest et al, 1969; Kirchner, 1966; Lawler, 1967; Nealey & Owen, 1970; Porter & Lawler, 1968; Thornton, 1966; Webb, 1953), a second level supervisor (Prien & Liske, 1962), or students (Borg & Hamilton, 1956.) In seven studies, including a few of the above, some degree of convergence was found. At least one significant correlation was found with supervisors' ratings (Borg & Hamilton, 1956; Heneman, 1974; Lopez, 1968; Prien & Liske, 1962; Thornton, 1966; Williams & Seller, 1973), peers (Borg & Hamilton, 1956), and students (Webb & Nolan, 1953). Three studies of work motivation investigated the relationship of effort and performance ratings. In all three, individuals and their supervisors converged on ratings of effort. Williams and Seller (1973) provided the best evidence of convergent validity for

the ratings of both performance and effort. These positive results may be due to clear conceptualization of the constructs and the development of good rating scales. Significant correlations were found for both global and dimensional ratings of the two constructs but there was more agreement on performance than effort ratings.

Discriminant Validity

It was difficult to form any firm conclusions regarding discriminant validity because of (a) small sample sizes, (b) incomplete analyses, (c) the difficulty of meeting all three of Campbell and Fiske's requirements, and (d) the relative lack of discrimination in ratings in general. Therefore, only tentative comments can be made in this area. Minimal evidence for discriminant validity of self-ratings was found. Three studies (Charest et al, 1969; Heneman, 1974; Williams & Seiler, 1973) found that only Campbell and Fiske's (1959) first two requirements were met to any extent. Serious problems with self-ratings were identified in the three studies cited above plus studies by Lawler (1967) and Nealey & Owen (1970).

Miscellaneous Findings

Charest et al (1969) gathered self, supervisor, and peer appraisals on two occasions separated by three weeks. Self-appraisals showed the least stability followed by supervisor and peer ratings.

Hamilton and Dickinson (1978) used ratings by incumbents, supervisors, and test experts in a synthetic validity study. Components

of J-coefficients included relationships of job elements and job performance and job elements and test performance. J-coefficients always correlated with actual validities when supervisors' data were used, but never correlated when incumbent self-ratings were used.

Three factor analytic studies were located. Klimoski and London (1974) and Prien and Liske (1962) both found a general halo factor plus rater-group factors. In the first study, Self, Supervisor, and Peer factors emerged; in the second, Self and Supervisor factors emerged. Thornton (1966) found substantially different factor structures for self-and supervisor-appraisals of executive performance. Only the supervisor's factor scores correlated with a criterion of promotability.

In summary, I conclude that self-appraisals tend to manifest more leniency, less variability, less halo, less agreement with other sources, less discriminant validity, and less reliability than ratings by supervisors and peers. Adequate contruct validity was found in a few studies.

Discussion

The inescapable conclusion is that individuals have a significantly different view of their own job performance than that held by other people. This discussion will explore some of the implications of the specific findings for various uses of self-appraisals.

At the beginning of the paper several purposes of self-appraisals were listed. We can now explore some implications of each psychometric characteristic of self-appraisals for these purposes.

Performance evaluation. The psychometric characteristics of self-appraisals may present a number of problems for their administrative use. Inflated evaluations and low correspondence with the supervisor's appraisal may make the performance appraisal discussion difficult for both parties. Different factor structures imply different perceptions of role requirements and may lead to "we/they" feelings. Written appraisals and completion of performance ratings forms will accentuate differences and tend to rigidify positions. By implication, the individual should not be forced to take a firm stand that self appraisals require. A more idealistic view would be that the self-appraisals will make apparent any implied areas of disagreement and lead to healthy conflict over the evaluation. Low variability in the ratings means that individuals do not distinguish among themselves to a great extent. This may be a disadvantage in making administrative decisions such as distribution of scarce resources. Lower halo suggests that intra-individual strengths and and weaknesses are being disgnosed more completely.

Identification of training needs. The leniency effect would imply that certain training needs would go undetected. As a result

the individual may not get appropriate coaching or training programs. This problem would be further kindled by low correlation between self-and supervisors-ratings. The wrong individuals would be given training if we rely on self-ratings. Furthermore, low variability in self-ratings would make it appear that everyone needed the same types of training. Different factor structures might lead to inappropriate arrangement of the content of a training program--an arrangement which was meaningful to supervisors but not employees. A bright spot in these findings is that the low halo among self-appraisals implies a better diagnosis of each individuals' training needs. Such disgnosis is the basis of individualized developmental planning.

Criterion measurement. Self-appraisals might be used as criterion measures in test validation and evaluation of training programs. How would their character effect results? Elevation of means would not directly affect prediction-criterion correlations, but restriction in range in criteria may lower validity coefficients and may lead to the rejection of otherwise valid tests. With low variance and low validity the standard error of estimate will be lower. There are implications here for differential validity and discrimination among ethnic and sex subgroups who might provide self-ratings. Low halo may result in an apparent low internal consistency of sub-criteria. Lack of correlation between self-

and supervisor-appraisals could result in the finding that some tests are valid for one measure of performance and not another. This could lead to the conclusion that the test validity does not generalize. The use of self-appraisal factor scores may lead to improper weighting of sub-criteria. For the evaluation of training impact, inflated self-appraisals may lead to exaggerated claims for training effectiveness.

Measurement of constructs. The major inadequacy here is the lack of discriminant validity. A large method bias exists--which in this case is the self as a source of data. On the negative side this means that the individual rater is a large source of variance in all ratings of various performance constructs. In experimental studies, leniency may lead to overstatement of the effects of treatment variables; in correlational studies, low variability may limit validity coefficients.

On the positive side, the perception of individuals of their own job performance is an important bit of phenomonological data. Lack of correlation between self- and supervisor-appraisals and different factor structures further substantiates that separate performance domains exist and should be explored. There are several areas where self-appraisals are the most appropriate data for investigation. In expectancy theory, a worker's cognitions of performance are more directly a function of motivational force.

Cognitions are an intervening variable between motivational force and objective performance and should be studied. Equity theory (Adams, 1965) explains that individuals seek a balance of their own contributions and outputs in comparison with other people. Imbalance or inequity causes dissonance which may be reduced by altering the quality or quantity of work until the individual sees a balance. Again, it is the cognition of performance which must be studied. According to Korman (1970), self-esteem is composed of chronic, task-specific and social self-esteem. Task-specific self-esteem can be operationalized as self perceptions of job effectiveness.

Selection. When used for gathering information at time of selection, self-appraisals are likely to lead to inflated statements of qualification and little differentiation among applicants. On the other hand, they could tap unique sources of information about the applicant. If found valid, they would contribute uniquely to the prediction of job performance in conjunction with other predictors. The relative lack of halo would provide better diagnosis of an applicant's individual strengths and thus provide a foundation for better placement decisions.

At the risk of over-simplification we can react to these findings in two ways:

(1) Negative view - self-appraisals are psychometrically inadequate and of limited value. They should be used with great caution and not for personnel decision-making.

(2) Positive view - self-appraisals are psychometrically different from appraisals by others; they provide unique information about job performance that is potentially as valid as other information; in fact, for some research areas the individual's phenomonological appraisal is the only appropriate data to use.

My own view is somewhere in between. Self-appraisals should be used quite carefully. Be aware of potential dangers. The psychometric characteristics of self-appraisals may effect research results. Supervisors using self-appraisal with subordinants should be warned that their own appraisals and the self-appraisals of their subordinates might be quite discrepant.

What explains the findings of this body of literature? The present review does not provide many clues. The observed psychometric qualities of self-appraisals does not appear to be a function of scale format, amount of rater training, type of judgment, or purpose of appraisal. More research is needed. Theoretical and empirical investigations would benefit from inclusion of some concepts in cognitive psychology (Bandura 1977, 1978) and self theory (Markus, 1977; Zander, 1963).

References

Adams, J. S. Inequity in social exchange. In L. Berkowitz (ed.) Advances in Experimental Social Psychology (Vol. 2). New York: Academic Press, 1965.

Baird, L. S. Self and superior ratings of performance: As related to self-esteem and satisfaction with supervision. Academy of Management Journal, 1977, 20, 291-300.

Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. Psychological Bulletin, 1977, 84, 191-215.

Bandura, A. The self system in reciprocal determinism. American Psychologist, 1978, 33, 344-358.

Bassett, G. A. & Meyer, H. H. Performance appraisal based on self-review. Personnel Psychology, 1968, 21 , 421-430.

Borg, W. R. & Hamilton, E. R. Comparison between a performance test and criteria of instructor effectiveness. Psychological Reports, 1956, 2, 111-116.

Campbell, D. T. & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Campbell, J. P. & Pritchard, R. D. Motivation theory in industrial and organizational psychology. In Dunnette, M. D. (Ed.) Handbook of Industrial and organizational psychology, Chicago: Rand McNally, 1976.

Charest, A. G., Cowart, D. G. & Goodman, P. S., Multi-instrument, multi-rater, multi-trait method for assessing measures of managerial performance, American Psychological Association, Experimental Publication System, 3, Ms. No. 092A, December, 1969.

Hall, H. L. An evaluation of the upward mobility assessment center for the Bureau of Engraving and Printing. U. S. Civil Service Commission, PB-262-521, July, 1976.

Hamilton, J. W. & Dickinson, T. L. Validity of policy-capturing as a procedure for synthetic validation. Paper presented at the meeting of the Rocky Mountain Psychological Association, Denver, April, 1978.

Heneman, H. G. Comparisons of self- and superior-ratings of managerial performance. Journal of Applied Psychology, 1974, 59, 638-642.

Kirchner, W. K. Relationships between supervisory and subordinate ratings for technical personnel. Journal of Industrial Psychology, 1966, 3, 57-60.

Klimoski, R. J. & London, M. Role of the rater in performance appraisal. Journal of Applied Psychology, 1974, 59, 445-451.

Korman, A. K. Toward a hypothesis of work behavior. Journal of Applied Psychology, 1970, 54, 31-41.

Lawler, E. E., III The Multitrait-multirater approach to measuring managerial job performance. Journal of Applied Psychology, 1967, 51 , 369-380.

Markus, H. Self-schemata and processing information about the self. Journal of Personality & Social Psychology, 1977, 35, 63-78.

McGregor, D. An uneasy look at performance appraisal. Harvard Business Review, 1957, 35(3), 89-94.

McGregor, D. The human side of enterprise. New York: McGraw-Hill, 1959.

Mendel, R. M. & Dickinson, T. L. The motivational determinants of task performance in a non-industrial milieu; A modification and extension of Vroom's model. Paper presented at the meeting of the Midwestern Psychological Association, Detroit, May, 1971.

Mendel, R. M. & Dickinson, T. L. A critical evaluation of the dependent variable in expectancy theory: Cognitive versus objective performance. Unpublished manuscript, undated.

Meyer, H. H. The annual performance review discussion--making it constructive. Personnel Journal, 1977, 56, 508-511.

Meyer, H. H., Kay, E., & French, J. R. P., Jr. Split roles in performance appraisal. Harvard Business Review, 1965, 63(1), 123-129.

Miner, J. Management appraisal: A capsule review and current reference. Business Horizons, 1968, 11(5), 83-96.

Mitchell, T. R. & Albright, D. W. Expectancy theory predictions of the satisfaction, effort, performance, and retention of Naval aviation officers. Organizational Behavior and Human Performance, 1972, 8, 1-20.

Nealey, S. M. & Owen T. W. A multitrait-multimethod analysis of predictors and criteria of nursing performance. Organizational Behavior & Human Performance. 1970, 5, 348-365.

Parker, J. W., Taylor, E. K., Barrett, R. S. & Martens, L. Rating scale content: III. Relationship between supervisory- and self-ratings. Personnel Psychology, 1959, 12, 49-63.

Peters, R. E. & Campbell, J. T. Diagnosis of Training Needs of B-24 Mechanics from Supervisory Ratings and Self-Ratings. Personnel Research Laboratory, Air Force Personnel and Training Research Center, April, 1955.

Porter, L. W. & Lawler, E. E., III Managerial attitudes and performance. Homewood, Ill.: Richard D. Irwin, Inc., 1968.

Prien, E. P. & Liske, R. E. Assessments of higher-level personnel: III. Rating criteria: A comparative analysis of supervisory ratings and incumbant self-ratings of job performance. Personnel Psychology, 1962, 15, 187-194.

Schwartz, D. J. A job sampling approach to merit system examining. Personnel Psychology, 1977, 30, 175-185.

Thornton, G. C., III The relationship between supervisory and self-appraisals of executive performance. Unpublished doctoral dissertation, Purdue University, 1966.

Thornton, G. C., III The relationship between supervisory- and self-appraisals of executive performance. Personnel Psychology, 1968, 21, 441-455.

Waldman, D. A. & Thornton, G. C., III A comparison of supervisors' self-appraisals and their administrators' appraisals. Journal of Medical Group Practice Administrator, in press.

Webb, W. B. & Nolan, C. Y. Ratings of instructional proficiency: Student ratings, self-ratings, and supervisor-ratings. St. Louis, Mo.: Washington University, Office of Naval Research Contract, NONR 816 (02), 1953. Cited in Borg, W. K. & Hamilton, E. R. Comparison between a performance test and criteria of instructor effectiveness. Psychological Bulletin, 1959, 56, 81-105.

Whisler, T. L. & Harper, S. T., Performance appraisal, Holt, Rinehart, & Winston, New York, 1962.

Wickert, F. R. & McFarland, D. E. (Eds.) Measuring executive effectiveness. New York: Appleton-Century Crafts, 1967.

Williams, W. E. & Seiler, D. A. Relationship between measures of effort and job performance. Journal of Applied Psychology, 1973, 57, 49-54.

Zander, A. F. (Ed.) Performance appraisals: Effects on employees and their performance. Ann Arbor, Mich: Foundation for Research on Human Behavior, 1963.